# UC Irvine
## UC Irvine Previously Published Works

**Title**

Understanding Disorder Within Variation: Production of English Grammatical Forms by English Language Learners

**Permalink**

https://escholarship.org/uc/item/0xn0z0qt

**Journal**

Language Speech and Hearing Services in Schools, 49(2)

**ISSN**

0161-1461

**Authors**

Bedore, Lisa M
Peña, Elizabeth D
Anaya, Jissel B
et al.

**Publication Date**

2018-04-05

**DOI**

10.1044/2017_lshss-17-0027

Peer reviewed

**Research Article**

# Understanding Disorder Within Variation: Production of English Grammatical Forms by English Language Learners

Lisa M. Bedore,[a] Elizabeth D. Peña,[b] Jissel B. Anaya,[a] Ricardo Nieto,[a] Mirza J. Lugo-Neris,[a] and Alisa Baron[a]

**Purpose:** This study examines English performance on a set of 11 grammatical forms in Spanish–English bilingual, school-age children in order to understand how item difficulty of grammatical constructions helps correctly classify language impairment (LI) from expected variability in second language acquisition when taking into account linguistic experience and exposure.
**Method:** Three hundred seventy-eight children's scores on the Bilingual English–Spanish Assessment–Middle Extension (Peña, Bedore, Gutiérrez-Clellen, Iglesias, & Goldstein, 2008) morphosyntax cloze task were analyzed by bilingual experience groups (high Spanish experience, balanced English–Spanish experience, high English experience, ability (typically developing [TD] vs. LI), and grammatical form. Classification accuracy was calculated for the forms that best differentiated TD and LI groups.

**Results:** Children with LI scored lower than TD children across all bilingual experience groups. There were differences by grammatical form across bilingual experience and ability groups. Children from high English experience and balanced English–Spanish experience groups could be accurately classified on the basis of all the English grammatical forms tested except for prepositions. For bilinguals with high Spanish experience, it was possible to rule out LI on the basis of grammatical production but not rule in LI.
**Conclusions:** It is possible to accurately identify LI in English language learners once they use English 40% of the time or more. However, for children with high Spanish experience, more information about development and patterns of impairment is needed to positively identify LI.

S
peech-language pathologists (SLPs) are charged with identifying bilingual children with language impairments (LIs). In recent years, there has been an increase in the availability of valid and reliable measures, yet bilingual children continue to be both overidentified and underidentified with language disorders (Sullivan & Bal, 2013). One reason for misidentification can partially be attributed to the false notion that there is a strict dichotomy between language difference and disorder. Because funding for special education and English language learning support comes from different sources, in some school districts, children are routinely channeled to different programs to avoid "double dipping" or providing special education and English language services. This is unfortunate because children with disabilities who are English language learners (ELLs) are entitled to appropriate educational services, including English language services and special education services (Lhamon & Gupta, 2015). With the goal of reducing misidentification of ELLs, SLPs are tasked with determining whether differences in children's English language performance are due to LI or errors associated with acquiring English as a second language. We know that bilingual children in the United States have heterogeneous levels of experiences and use of each language, and we expect that this will lead to variability in their performance on language tasks. Thus, we need to systematically profile and identify LI within the expected variability in ELLs' performance.

It is the special nature of grammatical difficulties that makes grammar a relatively robust clinical marker of LI in monolingual English speakers (Bedore & Leonard,

[a]Department of Communication Sciences & Disorders, The University of Texas at Austin
[b]School of Education, University of California, Irvine

Correspondence to Lisa M. Bedore: lisa.bedore@austin.utexas.edu

1998; Rice & Wexler, 1996). Bilingual learners in the process of acquiring English as a second language demonstrate difficulties with many of the same grammatical forms that are considered to be clinical markers of LI (Gutierrez-Clellen & Simon-Cereijido, 2007; Paradis, 2008). A challenge for clinicians is that they must determine if the grammatical difficulties that they observe in the language of a bilingual child are within the expected range of performance or if they are indicative of language learning difficulties. Because clinicians lack information about normal variability and how this might differ from LI, bilingual children are often overreferred to special education or services are delayed (Bedore & Peña, 2008). The focus of the current study is to determine if children with and without LI can be accurately classified with measures of grammatical structures that show variable production in bilingual language learners.

## Grammatical Difficulties of Children With LI

English-speaking children with LI differ from their typically developing (TD) peers across a range of linguistic forms. Overall, their language is less complex relative to their peers as indexed by mean length of utterance (MLU; Rice et al., 2010), complexity of predicates (Johnston & Kamhi, 1984), and production of arguments (Grela & Leonard, 2000). They perform lower than their age-matched peers across most grammatical constructions that have been tested in the literature, including articles, definite and indefinite pronouns, possessives, plural markers, present and past tense marking, prepositions, and copulas and auxiliaries (see Leonard, 2014, for an overview). In addition, children with LI are less likely to produce morphosyntactic constructions that require insertion of "do," such as negation (e.g., "These children like cookies these children *do not*"), questions (e.g., "Do you like cookies?"), or movement of a *wh*-question word to the beginning of the utterance (e.g., "What kind of cookies do you like?"; Hewitt, Hammer, Yont, & Tomblin, 2005; Rice, Hoffman, & Wexler, 2009).

The patterns described above illustrate the various grammatical forms children with LI have difficulties with, but the degree of difficulty these children have with these forms varies widely. For some grammatical forms, the differences between TD and LI groups are statistically significant but relatively small. For example, children with LI and their MLU-matched peers often score in the same range on irregular past tense (Leonard, Eyer, Bedore, & Grela, 1997; Rice, Wexler, Marquis, & Hershberger, 2000). On the whole, children with LI produce plural *–s* somewhat less accurately than do their TD MLU- and age-matched peers, but the performance of children with TD and LI can overlap. A subset of these forms, such as verb tense marking, demonstrates more robust differences, for example, 5-year-olds with LI score 60%–70% below their age-matched peers on these forms across studies (Leonard et al., 1997; Rice & Wexler, 1996). These forms are often referred to as *clinical markers of language impairment* because they maximally differentiate the production patterns of children with and without LI and are thus informative in the diagnostic context. English clinical markers of LI include third-person singular present tense *–s,* regular past tense *–ed,* the auxiliary *do,* and the copula and auxiliary *be* (Bedore & Leonard, 1998; Rice & Wexler, 1996).

## Grammatical Challenges of ELLs With TD and LI

Children acquiring English as a second language who have no identified risk for LI have difficulty with many of the same English grammatical constructions as the monolingual English-speaking children with LI described above. Forms that have been shown to be more difficult for ELLs relative to their monolingual peers include plural, possessives, pronouns, progressive forms, third-person singular present tense, irregular and regular past tense, and prepositions (Chondrogianni & Marinis, 2011; Marinis & Chondrogianni, 2010; Nicholls, Eadie, & Reilly, 2011; Paradis, 2016; Paradis, Emmerzael, & Sorenson Duncan, 2010; Taliancich-Klinger, Bedore, & Peña, 2017). However, the gaps in performance between TD bilingual children and their monolingual peers are not equal for all forms, nor does it take bilingual children equally as long to catch up to their monolingual peers in regard to all of these forms. For example, the English copula and auxiliary *be* forms seem much less difficult to learn than verb tense–related morphemes, such as past tense *–ed* or the third-person singular present tense *–s* (e.g., Nicholls et al., 2011; Paradis, 2016).

It is well documented that experiential factors partially account for the level of grammatical knowledge in bilingual learners. Among the factors demonstrated to affect second language knowledge are current and cumulative exposure to the second language (L2) and quality and sources of input (Bedore et al., 2012; Hammer et al., 2012; Paradis, 2017). Bedore et al. (2012) found that current language use accounted for 62% of the variability in preschool-age children's dominance scores (difference between Spanish and English scores) in the area of morphosyntax on an experimental version of Bilingual English–Spanish Oral Screener (BESOS; Peña, Bedore, Gutiérrez-Clellen, Iglesias, & Goldstein, 2006). Focusing specifically on the production of third-person present tense and past tense marking (on the basis of the Rice/Wexler Test of Early Grammatical Impairment; Rice & Wexler, 2001), both Chondrogianni and Marinis (2011) and Blom and Paradis (2015) reported moderate correlations between production accuracy and measures of cumulative English exposure. Morphosyntactic production, as tested on the Diagnostic Evaluation of Language Variation (Seymour, Roeper, & de Villiers, 2003), was also strongly correlated with experience in Chondrogianni and Marinis (2011). These findings highlight the importance of accounting for the role of language experience when considering expected levels of language knowledge for bilingual children who are in the process of acquiring their second language.

Given the variability observed in TD bilingual children on the grammatical forms discussed above, the key question is whether the performance of bilingual children with LI can be differentiated from their TD bilingual peers. One of the earliest studies to address this question is that of Jacobson

and Schwartz (2005), who focused on the extent to which past tense production differentiated between relatively balanced Spanish–English bilingual children with and without LI. Based on elicited productions, TD children were more likely to produce overregularization errors (e.g., *runned* for *ran*) as opposed to their LI peers who produced more omissions (e.g., *walk* for *walked*). Jacobson and Schwartz classified these children on the basis of regular past tense production with 89.5% accuracy and irregular past tense production with 82.5% accuracy. There are reliable differences between children with and without LI for bilinguals from different language backgrounds who are in the process of acquiring English for tense marking forms. For example, Blom and Paradis (2015) documented an average difference of 30% in production accuracy between children with TD and LI for 5- to 7-year-olds who had been learning English for about 3.5 years on average. Verhoeven, Steenge, and van Balkom (2012) documented similar patterns of performance on verb tense marking for children acquiring Dutch as a second language. These findings confirm the difficulty bilingual children with LI have with tense marking as documented for monolingual English speakers. However, there is considerable overlap between the performance of children with and without LI, and discriminant analyses were not included as part of these studies. In a broader look at bilingual children with and without LI, Gutierrez-Clellen and Simon-Cereijido (2007) tested grammatical performance on a range of forms (possessive –*s,* plural, third person present tense, regular and irregular past tense, copula, auxiliary, negation, relative clause, and adverbial clauses) using cloze and sentence repetition tasks. On the basis of a composite score, they demonstrated between-groups differences and classified balanced bilinguals and English-dominant speakers with approximately 80% sensitivity (correct classification of children with LI) and 90% specificity (correct classification of children without LI). These findings highlight the potential utility of a grammatical marker approach to identifying bilingual children, but it may be useful to consider forms beyond tense markers and to examine the performance of ELLs in earlier stages of English acquisition.

### Summary and Questions

It is challenging to identify LI in the face of the variability observed in bilingual children who are in the process of acquiring English. There are at least two sources of variability that should be considered when identifying LI within the context of language difference. First and foremost, the grammatical constructions that should be the focus of these efforts should be considered. Specifically, tense markers, which have been shown to be reliable markers for young English-speaking children with LI, are also challenging for ELLs who are in the process of learning English. Other grammatical forms are also quite difficult for ELLs with LI and may prove to be useful as clinical markers. Thus, it is important to systematically test those forms that are known to be good clinical markers for English learners and other forms that are known to be challenging for ELLs with

LI. Second, it is important to consider the extent to which experience with the L2 influences classification accuracy. Language experience, calculated on the basis of language use, accounts for a significant amount of the variance in children's performance on morphosyntax measures. With these findings in mind, we address the three questions listed below in an effort to understand how we can most effectively identify LI in the context of the variability that is typically observed in bilingual (Spanish–English-speaking) children who are still acquiring their second language.

1.  Do level of bilingual experience and ability influence children's performance on English morphosyntactic forms as measured by performance on the experimental version of the cloze subtest of the Bilingual English–Spanish Assessment–Middle Extension (BESA-ME; Peña, Bedore, Gutiérrez-Clellen, Iglesias, & Goldstein, 2008)?

2.  Are there differences in performance across English morphosyntactic forms as a function of children's language ability and bilingual experience group (high Spanish experience [HSE], balanced English–Spanish experience [BESE], and high English experience [HEE])?

3.  Does classification accuracy of English morphosyntax vary as a function of bilingual experience group?

## Method
### Participants

Data for the current study were selected from three existing data sets, which included Spanish–English bilingual children with different levels of bilingual experience (Peña, Bedore, & Gillam, 2006; Peña, Bedore, & Griffin, 2010). In these larger studies, participants were recruited from schools that enroll high numbers of bilingual Latino students in school districts in Texas, Colorado, and Utah. Participants were selected for the current analyses if they spoke English and Spanish, were less than 10 years old, had completed the English cloze task of the experimental version BESA-ME (Peña, Bedore, Gutiérrez-Clellen, et al., 2006), and had sufficient data for bilingual level and ability status determination.

A total of 378 Spanish–English bilingual children (313 with typical development and 65 with LI) between the ages of 7 and 10 years were included in this study. Participants included 184 (37 with LI) from the Phenotype Assessment Tools for Bilingual (Spanish–English) Children (Peña & Bedore, 2006); 35 (two with LI) children from Diagnostic Markers of Language Impairment (Peña et al., 2006); and 159 children (26 with LI) from the Cross-Language Outcomes of Typical and Atypical Development in Bilinguals (Peña et al., 2010). Across the three studies, parents completed questionnaires to determine current exposure to English and Spanish, age of first exposure to English, and demographic information. Children included in the current study used English and Spanish at least 20% of the time. The primary indicator of socioeconomic status was mother education,

which was collected during parent interviews and calculated on the basis of the Hollingshead Four-Factor Index of Social Status (Hollingshead, 1975). To quantify the children's language skills, parents and teachers rated language knowledge in both languages and children completed language testing in English and Spanish. The breakdown of age, age of first exposure to English, sex, socioeconomic status, semantics subtest standard scores in English and Spanish, and parent ratings of English and Spanish language skills for all participants is shown in Table 1.

## Ability Status

Across the three studies, participants were assigned to TD or LI groups on the basis of indicators of LI, including parent and teacher questionnaires, semantics, morphosyntax, narratives, screening data, and/or SLP clinical judgment. Parent and teacher indicators were derived from questionnaires that included questions about language development in vocabulary, sentence length, grammar, comprehension, and articulation. Semantics and morphosyntax indicators were based on scores on the BESA (Peña et al., 2018) or the field test version of the BESA-ME (Peña, Bedore, Gutiérrez-Clellen, Iglesias, & Goldstein, 2016) and the Test of Oral Language Development–Third Edition (Newcomer & Hammill, 1997). A narrative indicator included information from a narrative sample on the basis of scores on the English Test of Narrative Language (TNL-E; Gillam & Pearson, 2004), the Test of Narrative Language Spanish Experimental Version (TNL-S; Gillam, Peña, Bedore, & Pearson, in development), or narrative retells and tells in Spanish and English on the basis of a wordless picture book. The screening indicator was based on scores from the BESOS (Peña, Bedore, Gutiérrez-Clellen, Iglesias, & Goldstein, in development). SLP ratings were either referral to the study as LI by an SLP or clinical judgment of the presence of LI by an SLP with expertise in bilingualism. In the three studies from which we drew participants, these indicators were used in slightly different combinations to verify language ability status. Table 2 lists the specific measures and decision rules used to determine LI for children from each study.

Children on the Phenotype Assessment Tools for Bilingual (Spanish–English) Children study were identified with LI if they met two of the three indicators: (a) parent or teacher rating on Instrument to Assess Language Knowledge (ITALK) below 4.2 of 5 in both languages; (b) TNL in both languages more than 1 $SD$ below normative mean; and (c) identification by a school-based SLP as having LI. On the Diagnostic Markers of Language Impairment study, children were identified with LI on the basis of three expert SLPs who rated semantics, morphosyntax, and narratives in Spanish and English on the basis of test responses and transcribed narrative samples collected when children were in the first grade. Overall judgment of impairment was based on a 6-point scale (0 = *severe/profound,* 1 = *moderate,* 2 = *mild,* 3 = *low normal,* 4 = *normal,* and 5 = *above normal*). Children were identified with LI if two of the three raters assigned a rating of 2 or less. Note that this group

of children was identified using indicators from the BESA when they were in kindergarten and first grade and tested with the BESA-ME (Peña et al., 2016) when they were in third grade. On the Cross-Language Outcomes of Typical and Atypical Development in Bilinguals study, children were identified with LI if they met four of the five criteria: (a) Parent or teacher rating below 4.2 (of 5) in both languages; (b) BESA-ME field test version, morphosyntax more than −1 $SD$ from the normative mean in both languages; (c) BESA-ME field test version, semantics more than −1 $SD$ from the normative mean in both languages; (d) BESOS composite more than −1 $SD$ from the normative mean in both languages administered 1 year prior to testing; and (e) TNL more than −1 $SD$ from the mean in both languages.

### Bilingual Experience Grouping

Participants were assigned to one of three bilingual experience groups on the basis of parent and teacher questionnaires of language use and exposure at home and at school (Peña et al., 2018). Children were considered HEE if they used Spanish 20%–40% of the time, BESE if they used Spanish 41%–59% of the time, and HSE if they used Spanish 60%–80% of the time. Note that English use is the inverse of Spanish for these children.

## Measures

### Parent and Teacher Interviews

Parent and teacher interviews were conducted by phone using the Bilingual Input–Output Survey (Peña et al., 2018) to obtain information on the children's cumulative exposure to each language since birth. Parents reported on the language of the home and language of the school or day care on a year-by-year basis. Parents and teachers provided information on children's language input and output on an hour-by-hour basis for a typical weekday and weekend day at home and typical weekday at school (Gutiérrez-Clellen & Kreiter, 2003; Peña et al., 2018). These data were projected for a full 7-day week to generate an average percentage of input and output in each language. This average was the basis for the bilingual experience groupings.

Parents and teachers also responded to questions for each language using the ITALK (Peña et al., 2018). Parents, for example, rated their child's vocabulary, speech intelligibility, sentence length, grammatical accuracy, and language comprehension in each language on a 5-point scale (e.g., vocabulary rating ranged from *a few words* (1) to *extensive vocabulary* (5)). The five scores are averaged for each language to determine if concern exists. The ITALK has a reported sensitivity and specificity of 0.80 when applying a cut score of 4.18 in both languages.

### BESOS

The BESOS (Peña et al., in development) assesses semantics and morphosyntax performance in order to identify children with possible LI through subtests in English and Spanish. There are age/grade-specific versions for children between prekindergarten and third grade. The

**Table 1.** Participant information presented in means and standard deviations.

| Measure | HEE | | BESE | | HSE | | All | |
|---|---|---|---|---|---|---|---|---|
| Status | TD | LI | TD | LI | TD | LI | TD | LI |
| N | 70 | 10 | 129 | 24 | 114 | 31 | 313 | 65 |
| Sex | 39 F, 31 M | 3 F, 7 M | 64 F, 65 M | 9 F, 15 M | 60 F, 54 M | 8 F, 23 M | 163 F, 150 M | 20 F, 45 M |
| Age in months | 104.80 (9.95) | 105.60 (10.55) | 101.02 (10.16) | 98.92 (9.82) | 98.92 (8.44) | 97.58 (10.57) | 101.10 (9.74) | 99.31 (10.51) |
| SES[a] | 3.68 (1.71) | 3.00 (1.63) | 2.64 (1.50) | 2.43 (1.47) | 2.57 (1.56) | 2.81 (1.74) | 2.84 (1.63) | 2.70 (1.62) |
| Age of first exposure in years[b] | 2.01 (2.24) | 2.60 (2.22) | 2.42 (1.99) | 3.05 (2.30) | 3.68 (1.71) | 3.90 (1.68) | 2.79 (2.07) | 3.40 (2.04) |
| English Semantics SS | 97.17 (12.91) | 63.70 (18.49) | 92.90 (16.18) | 62.44 (21.28) | 88.03 (18.44) | 51.42 (18.20) | 92.10 (16.71) | 57.38 (19.98) |
| Spanish Semantics SS | 79.17 (25.08) | 67.90 (17.60) | 96.80 (13.89) | 60.37 (22.74) | 97.85 (11.96) | 59.49 (19.24) | 93.54 (17.79) | 60.89 (20.33) |
| English vocabulary[c] | 4.24 (0.92) | 3.67 (1.86) | 3.45 (1.34) | 2.48 (1.08) | 3.01 (1.23) | 1.86 (1.22) | 3.48 (1.29) | 2.36 (1.27) |
| English sentence length[d] | 4.54 (0.82) | 4.11 (0.93) | 4.25 (1.09) | 2.45 (1.30) | 3.32 (1.27) | 2.33 (1.14) | 3.99 (1.21) | 2.66 (1.32) |
| English grammar[e] | 4.28 (0.72) | 3.50 (0.76) | 3.63 (0.95) | 2.52 (1.03) | 2.87 (1.07) | 2.46 (1.10) | 3.53 (1.09) | 2.64 (1.08) |
| Spanish vocabulary[f] | 2.97 (1.78) | 2.50 (2.12) | 4.47 (0.88) | 3.30 (1.02) | 4.36 (1.03) | 3.77 (1.10) | 4.12 (1.31) | 3.40 (1.34) |
| Spanish sentence length[g] | 3.13 (1.99) | 3.10 (2.28) | 4.77 (0.60) | 3.87 (1.22) | 4.72 (0.82) | 4.00 (1.11) | 4.41 (1.28) | 3.81 (1.40) |
| Spanish grammar[h] | 2.87 (1.89) | 2.20 (1.75) | 4.43 (0.70) | 3.33 (0.87) | 4.24 (0.98) | 3.63 (0.89) | 4.03 (1.29) | 3.30 (1.15) |

*Note.* HEE = high English experience; BESE = balanced English–Spanish experience; HSE = high Spanish experience; TD = typically developing; LI = language impaired; F = female; M = male; SES = socioeconomic status Hollingshead Score; SS = standard score.

[a]6 missing. [b]5 missing. [c]25 missing. [d]29 missing. [e]34 missing. [f]31 missing. [g]33 missing. [h]36 missing.

**Table 2.** Overview of types of indicators and decision rules for the identification of LI by study.

| Indicators | Phenotypes | Diagnostic markers | Cross-language outcomes |
|---|---|---|---|
| Questionnaires | X | X | X |
| Semantics | | X | X |
| Morphosyntax | | X | X |
| Narratives | X | X | X |
| Screening | | | X |
| SLP rating | X | X | |
| Decision rule | 2 of 3 indicators from ITALK, TNL, and SLP ratings. | Rating of 2 or less on a 6-point scale by 2 of 3 raters. | 4 of 5 indicators from ITALK, BESA or BESA-ME and BESOS, and TNL |

*Note.* LI = language impairment; Phenotypes = Phenotype Assessment Tools for Bilingual (Spanish–English) Children; Diagnostic markers = Diagnostic Markers of Language Impairment; Cross-language outcomes = Cross-Language Outcomes of Typical and Atypical Development in Bilinguals; SLP = speech-language pathologist; ITALK = Instrument to Assess Language Knowledge; TNL = Test of Narrative Language; BESA = Bilingual English–Spanish Assessment; BESA-ME = Bilingual English–Spanish Assessment–Middle Extension (2016 field test version); BESOS = Bilingual English–Spanish Oral Screener.

semantics subtests contain items that assessed knowledge of categories or concepts. The morphosyntax subtests include cloze and sentence repetition items that targeted challenging forms in each language (e.g., past tense *–ed,* third-person present tense *–s,* and copulas in English and articles, direct object clitics, and subjunctive in Spanish). Standard scores (*M* = 100; *SD* = 15) were calculated. Previous analyses demonstrate that the preschool–kindergarten version of the BESOS has 90% concurrent sensitivity and 91% concurrent specificity and 95.2% predictive sensitivity and 71.4% predictive specificity from preschool to first grade using a cut score of −1 *SD* below the mean in the best language (Lugo-Neris, Peña, Bedore, & Gillam, 2015). Preliminary analysis from the first grade BESOS demonstrates that it has a sensitivity of 93% and specificity of 92% using a cut score of −1 *SD* below the mean in the best language. Similarly, analysis of the third grade BESOS shows a sensitivity of 80% and specificity of 94% using a cut score of −1 *SD* below the mean in the individual child's best language.

## BESA

The BESA (Peña et al., 2018) is a measure designed for and normed with Spanish–English bilingual children ages 4–6;11 (years;months) in the United States. It has pragmatics, phonology, semantics, and morphosyntax subtests in English and Spanish. For the purpose of classifying participants who were employed in the current analyses, we used only the semantics and morphosyntax subtests. Semantics items focus on common home and school concepts. Morphosyntax includes cloze items and sentence repetitions. Standard scores with a mean of 100 and *SD* of 15 are derived for each subtest. Sensitivity for the Semantics subtests ranged 80%–83% for English and 72%–89% for Spanish. Specificity for the Semantics subtests ranged 78%–86% for English and 78%–88% for Spanish. Sensitivity for the morphosyntax subtests ranged 87%–89% for English and 78%–91% for Spanish. Specificity for the morphosyntax subtests ranged 81%–88% for English and 81%–88% for Spanish. Coefficient alpha for the Semantics is .89 for English

and .87 for Spanish. Coefficient alpha for the morphosyntax is .97 for English and .95 for Spanish.

## BESA-ME

The BESA-ME (Peña et al., 2008, 2016) is an experimental measure that assesses language ability in U.S. bilingual children between 7–9;11 across semantics and morphosyntax domains in both English and Spanish. The experimental test version of the BESA-ME (Peña et al., 2008) was employed in the study but scoring for identification of LI was completed using a subset of items that make up the field test version (Peña et al., 2016). Specifically, the experimental version is composed of all the items in the second iteration of the test (which is a subset of the initial pilot version), whereas the field test version includes a subset of items that were most sensitive to impairment. All children completed the English and Spanish experimental version of the BESA-ME (Peña et al., 2008), and the data from the cloze task were employed as the outcome variable in this study. Standard scores with a mean of 100 and *SD* of 15 are derived for each of the BESA-ME subtests.

The field test version of the BESA-ME (Peña et al., 2016) was used as one of the indicators of ability for 7- to 10-year-old participants from the Cross-Language Outcomes of Typical and Atypical Development in Bilinguals. For the semantics subtests, items focused on expressive and receptive semantic knowledge, and expressive responses were permitted in English or Spanish. The morphosyntax subtests include grammatical cloze items and sentence repetitions that target difficult structures in each language. Preliminary data for the BESA-ME (Peña et al., 2016) field test version indicate that, for English semantics, sensitivity was between 69% and 76% and specificity was between 80% and 88% depending on age. English morphosyntax sensitivity ranged from 63% to 71% with specificity from 85% to 89% depending on age. The Spanish version of the semantics subtest has sensitivity ranging from 54% to 81% and specificity ranging from 87% to 89%, depending on age. Spanish morphosyntax has a sensitivity ranging from 80% to 91% and specificity from 92% to 97% depending

on age. Generally, whereas sensitivity on the semantics subtest was lower than expected for given ages, specificity was acceptable. Morphosyntax has generally higher sensitivity and specificity. This pattern of results indicates that children who score in the language-impaired range are likely to have impairment, but it is possible that some children are missed by a single subtest in a given language. Composites incorporating semantics and morphosyntax across the two languages are expected to be more stable indicators of impairment. Coefficient alpha, which is an indicator of internal stability, for Spanish morphosyntax was .89 for first grade and .87 for third grade. For English, morphosyntax coefficient alpha was .84 for first grade and .83 for third grade. Coefficient alpha for Spanish semantics was .82 for first grade and .62 for third grade. Finally, coefficient alpha for English semantics was .65 for first grade and .70 for third grade.

### Narratives

The TNL-E (Gillam & Pearson, 2004) and the TNL-S (Gillam et al., in preparation) tested narrative comprehension and production abilities. Each test included three narrative tasks. The first task was a story retell with no visuals, and the other two tasks were story formulations, which were elicited by a picture or sequence of pictures. The structures of the Spanish and English versions of the TNL were similar but contained different stories. Standard scores with a mean of 100 and *SD* of 15 were derived for the English and Spanish versions separately. Sensitivity and specificity on the TNL-S have been derived using pilot data, and results showed sensitivity from .80 to .85 and specificity from .74 to .81. For English speakers, the TNL-E manual indicates a sensitivity of .92 and specificity of .87.

Elicited narratives on the basis of Mercer Meyer's wordless picture books (Mayer, 1967, 1969, 1973, 1974) were also used. Children told two narratives in each language. First, they were asked to retell a story on the basis of a model using a version of the scripts provided in the Systematic Analysis of Language Transcripts program (Miller & Iglesias, 2012). Then, they were given the second book, shown each of the pages, and then were asked to tell the story. Children were reminded to use the target language if they switched to their other language. Back-channeling cues were used to encourage the children to continue their stories (see Peña et al., 2010, for additional details). Productivity measures—including number of different words, mean length utterance, and total number of words—were derived based on the average values across the retell and tell. Percent of grammatical utterances was also derived as an indicator of LI.

### Procedure

For the present analyses, data for grammatical morpheme use in English was derived from the 37 cloze items included in the BESA-ME English morphosyntax experimental version (Peña et al., 2008) across all three larger data sets (Peña et al., 2008, 2016). Recall that the 2016 field test version of the BESA-ME is composed of a subset of the items first tested as part of the 2008 experimental version.

Of the 37 English morphosyntax cloze items from the experimental version in the current analyses, 18 are also included in the 2016 field test version, which was used as part of the morphosyntax indicator of LI in the Cross-Language Outcomes of Typical and Atypical Development in Bilinguals study (in addition to English sentence repetition items and Spanish morphosyntax subtests, along with other indicators of semantics, narratives, and parent/teacher ratings).

Morphemes with at least two exemplars were analyzed. These included the following forms: copula (two), passives (three), negatives (two), plurals (three), third-person singular (four), question inversion (six), past tense –*ed* (four), possessive –*s* (three), relative clause (three), prepositions (two), and irregular past (four). These items were drawn from the set used for the BESA-ME (Peña et al., in development). Examples of the items are displayed in Table 3. Percent accuracy of each grammatical morpheme by bilingual experience group is represented in Table 4.

## Results

To explore potential differences in English morphosyntactic skills between children of different bilingual experience groups and ability status, we conducted a two-way analysis of variance (ANOVA) with the standard score calculated from the cloze items on the experimental English version of the morphosyntax BESA-ME (Peña et al., 2008) subtest as the dependent variable. Main effects and the interaction between bilingual experience group (HSE, BESE, HEE) and ability status (LI, TD) were explored. Results indicated significant main effects for ability status, $F(1) = 130.64$, $p < .001$, partial $\eta^2 = .259$, and language experience group, $F(2) = 22.09$, $p < .001$, partial $\eta^2 = .106$. Children with LI scored lower than their TD counterparts, with a mean difference ($M_\Delta$) of $M_\Delta = -35.30$ points. Post hoc analyses using a Bonferroni correction were conducted to explore all pairwise differences among experience groups. Results revealed significant differences between BESE and HSE ($M_\Delta = 7.80$), BESE and HEE ($M_\Delta = -19.01$), and HSE and HEE ($M_\Delta = -26.81$) at the .05 alpha level. Thus, HSE children scored the lowest, whereas HEE children scored the highest.

### Morphosyntactic Forms

The second aim of this study was to explore differential performance on specific morphosyntactic forms from the BESA-ME English morphosyntax subtest. Recall that the cloze items of the morphosyntax subtest consisted of 11 forms. Figure 1 presents a line graph that illustrates the observed mean proportion correct (*y*-axis) on the 11 forms (*x*-axis) by bilingual experience group (colored lines) and ability status (line type). From left to right, the forms are ordered in decreasing order of average difficulty (i.e., average proportion correct across the sample). In general, Figure 1 highlights that, with a few exceptions, the TD children showed a higher percentage of correct responses on nearly all forms tested. Children with HEE-LI showed a slightly higher number of correct responses on negative

| Target | Model | Prompt |
|---|---|---|
| Copula | In this pond there is one fish. | And in this pond there (**are** two fish). |
| Passive | This truck is hit by the car. | And this car (**was hit** by the truck). |
| Negative | These men have moustaches. | And these men (**don't have** moustaches). |
| Plural | This girl has an apple. | And here she has many (**apples**). |
| Third-person singular present | Every day these dogs drink water. | And here this dog does it too. Every day the dog (**drinks** water). |
| Question inversion | This girl is watching her friend open a birthday present. | Tell me what she says. (**What is it?**) |
| Past tense –ed | Today he is walking his dog. | And yesterday he did it too. Yesterday he (**walked** the dog). |
| Prepositions | Now we are going to say where the cats are. These cats are in the jar. | Now the cats are (**on the plate**). |
| Irregular past | Today she is eating a banana. | Yesterday she did it too. She (**ate** a/the banana). |

and regular past forms than the BESE-TD and HSE-TD children. For prepositional phrase items, children with HEE-LI outperformed all TD subgroups. Furthermore, comparing the TD children across the three bilingual experience groups, HSE-TD children tended to score the lowest for nearly all forms. For the children with LI, the HEE-LI subgroup produced more items correctly on all item types compared with children with BESE-LI and HSE-LI. With the exception of copula items, children with HSE-LI answered the least number of items correct. Although children with LI tended to answer fewer items within forms, it is worth noting that looking across forms, this group performed at least as well on the more difficult items than the TD children performed on the easier items. Overall, these findings illustrate that, although there is variability in ELLs response patterns, the points of variability are limited, and it is possible to identify forms where differences in performance are interpretable.

We conducted a mixed-model ANOVA with the 11 grammatical forms serving as the within-subject factor and the bilingual experience group and ability status serving as between-subjects factors. The dependent measures were

the percentage of correct responses on the items associated with each of the 11 forms for each participant.

Mauchly's test of sphericity was statistically significant at the .05 alpha level, and thus the assumption of sphericity was not met. We therefore report results using the Huynh–Feldt adjustment, which statistically corrects for violation of the sphericity assumption. All within-subject interactions were statistically significant, including the two-way interaction between target and bilingual experience group, $F(15.69, 41685) = 2.96$, $p < .001$, partial $\eta^2 = .016$, two-way interaction between target and ability status, $F(7.85, 41170.) = 5.85$, $p < .001$, partial $\eta^2 = .015$, and three-way interaction between the three predictors, $F(15.69, 38922) = 2.76$, $p < .001$, partial $\eta^2 = .015$.

To further explore the specific patterns of performance between ability and bilingual dominance within each target type, we carried out a series of univariate between-subjects ANOVAs using the percentage correct on each target as the dependent variable. The results of these 11 ANOVAs are reported in Table 5. For brevity, this table only includes results for the effects found to be statistically significant at the .05 alpha level. A significant main effect of ability status

**Table 4.** Grammatical form accuracy in means and standard deviations.

| | HEE | | BESE | | HSE | |
|---|---|---|---|---|---|---|
| Grammatical form | TD | LI | TD | LI | TD | LI |
| Copula | *92.14 (21.93)* | *80.00 (39.74)* | **88.46 (22.97)** | **39.58 (36.05)** | **84.65 (26.71)** | **53.33 (39.25)** |
| Passive | **92.14 (17.83)** | **52.50 (34.26)** | **84.81 (25.51)** | **32.29 (37.21)** | **69.96 (34.91)** | **25.00 (35.96)** |
| Negative | *93.57 (16.86)* | *90.00 (42.51)* | **78.46 (34.10)** | **31.25 (35.55)** | **58.77 (41.67)** | **21.67 (33.95)** |
| Plural noun | **91.43 (17.66)** | **43.33 (29.89)** | **69.23 (29.76)** | **37.50 (24.70)** | **56.73 (31.96)** | **34.44 (32.14)** |
| Third-person singular | **88.93 (18.86)** | **55.00 (32.85)** | **71.73 (28.02)** | **27.08 (34.51)** | 53.95 (36.22) | **22.50 (31.04)** |
| Question inversion | **74.76 (25.18)** | **35.00 (26.75)** | **62.56 (33.41)** | **18.06 (22.48)** | **56.73 (34.19)** | **14.44 (17.90)** |
| Regular past | *78.93 (29.99)* | *62.50 (35.85)* | **53.65 (36.64)** | **16.67 (30.99)** | <u>43.20 (39.81)</u> | <u>16.67 (26.53)</u> |
| Possessive | **88.10 (24.76)** | **50.00 (37.71)** | **60.26 (40.13)** | **19.44 (33.93)** | 43.86 (41.20) | 8.89 (21.32) |
| Relative clause | <u>78.10 (25.31)</u> | <u>53.33 (34.95)</u> | **66.15 (37.61)** | **19.44 (29.35)** | **58.77 (37.61)** | **10.00 (19.87)** |
| Prepositional phrase | *42.14 (38.67)* | *55.00 (32.51)* | <u>44.62 (38.84)</u> | <u>25.00 (29.49)</u> | <u>41.23 (38.93)</u> | <u>18.33 (27.80)</u> |
| Irregular past | **60.00 (28.36)** | **10.00 (13.79)** | <u>31.35 (30.86)</u> | <u>10.42 (19.39)</u> | <u>21.49 (27.40)</u> | <u>4.17 (9.48)</u> |

*Note.* Bold = significant differences above 30% for LI and TD within dominance group; underline = significant differences, but below 30% for LI and TD within dominance group; italic = differences not significant for LI and TD within dominance group. HEE = high English experience; BESE = balanced English–Spanish experience; HSE = high Spanish experience; TD = typically developing; LI = language impaired.

**Figure 1.** The percentage of items correctly answered by children with LI and TD children at three levels of bilingual experience (HSE, BESE, HEE) on the 11 forms comprising the experimental English morphosyntax BESA-ME subtest. Items are ordered by the average difficulty or proportion correct. LI = language impairment; TD = typically developing; HSE = high Spanish experience; BESE = balanced English–Spanish experience; HEE = high English experience; BESA-ME = Bilingual English–Spanish Assessment–Middle Extension.
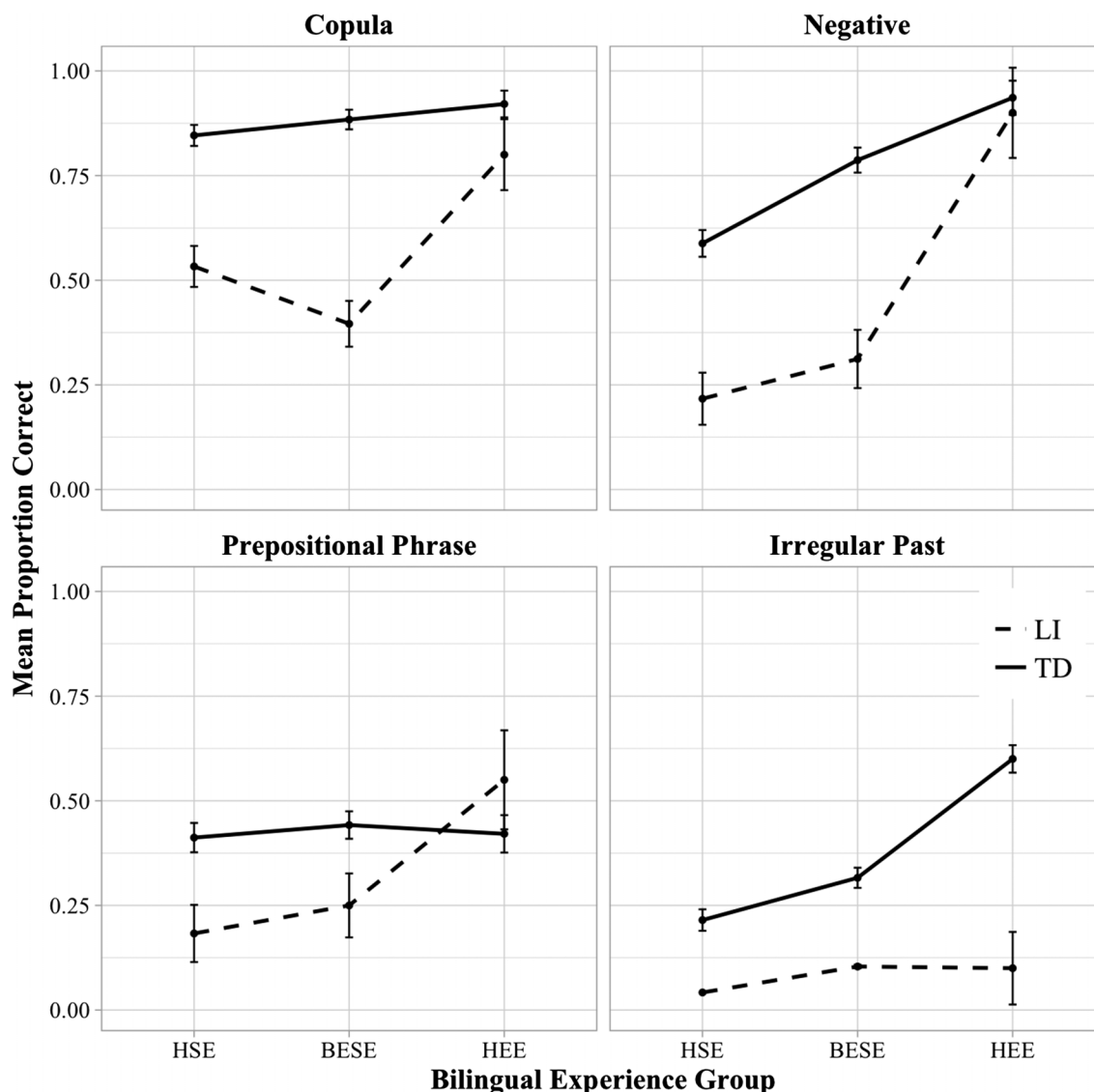


was found for all forms except prepositional phrase, with TD children having a higher percentage of correct scores on average in all cases. Results also revealed a significant main effect of bilingual experience group for all 11 forms. In each case, HEE children scored the highest number of items correct, and HSE children had the lowest scores. A significant interaction between bilingual experience group and ability status was observed for irregular past, prepositional phrase, copula, and negative forms. These interactions are displayed in Figure 2. As seen in this figure, although children with LI scored lower than TD children, the pattern of performance varied depending on their bilingual experience. Within the LI subgroup, HEE children outperformed both HSE and BESE children on copula, negative, and

propositional phrase forms. For irregular past items, the HEE-LI group answered fewer items correctly compared with children with BESE-LI, though we note the relatively large error observed for this estimate. The lowest performing LI group varied across forms. For copula items, children with BESE-LI answered the least number of items correctly. For negative, prepositional phrase, and irregular past items, children with HSE-LI produced fewer forms correctly. Comparing the general pattern of performance across bilingual experience groups within ability status (i.e., comparing the solid and broken lines within each plot in Figure 2), the greatest discrepancy between ability groups for copula and negative items was observed for BESE children, followed by HSE and HEE children. For prepositional

**Table 5.** Statistically significant results for the 11 ANOVAs examining the effects of BG, A, and their interaction on the proportion of items answered correctly.

| Target | Effect | $F$ | Partial $\eta^2$ |
|---|---|---|---|
| Copula | A | 57.72*** | .134 |
| | BG | 8.38*** | .043 |
| | A × BG | 6.13** | .032 |
| Passive | A | 102.56*** | .216 |
| | BG | 9.40*** | .048 |
| Negative | A | 32.33*** | .080 |
| | BG | 29.29*** | .136 |
| | A × BG | 5.06** | .026 |
| Plural noun | A | 60.99*** | .141 |
| | BG | 7.42** | .038 |
| Third-person singular | A | 64.69*** | .148 |
| | BG | 16.13*** | .080 |
| Question inversion | A | 82.39*** | .181 |
| | BG | 5.05** | .026 |
| Regular past | A | 24.51*** | .062 |
| | BG | 17.43*** | .086 |
| Possessive | A | 47.41*** | .113 |
| | BG | 17.59*** | .086 |
| Relative clause | A | 60.19*** | .139 |
| | BG | 10.75*** | .055 |
| Prepositional phrase | BG | 3.23* | .017 |
| | A × BG | 3.10* | .016 |
| Irregular past | A | 50.34*** | .119 |
| | BG | 8.50*** | .044 |
| | A × BG | 4.76** | .025 |

*Note.* ANOVA = analysis of variance; BG = bilingual group; A = ability status.

*$p < .05$. **$p < .01$. ***$p < .001$.

phrase items, the greatest discrepancy was observed for HSE children, followed by BESE and HEE. Interestingly, children with HEE-LI scored slightly higher than HEE-TD children, though we note the relatively large error for the HEE-LI estimate. For irregular past items, the greatest difference between ability groups was observed for HEE children, though children with LI and HSE-TD and BESE-TD children performed relatively poorly on this form in general. Tests of simple effects were conducted to determine which pair(s) of effects gave rise to the significant interactions. These follow-up analyses revealed (a) significant differences between HSE-LI/TD and BESE-LI/TD means ($M_\Delta = -.31$, $p < .001$ and $M_\Delta = -.49$, $p < .001$, respectively) on copula items and significant differences between HSE-LI/TD and BESE-LI/TD means on negative items ($M_\Delta = -.37$, $p < .001$ and $M_\Delta = -.47$, $p < .001$, respectively); (b) significant differences between BSE-LI/TD and BESE LI/TD means on prepositional phrase items ($M_\Delta = -.23$, $p < .01$ and $M_\Delta = -.20$, $p < .01$, respectively); and (c) significant differences between HSE-LI/TD, BESE-LI/TD, and HEE-LI/TD means on irregular past items ($M_\Delta = -.17$, $p < .01$, $M_\Delta = -.21$, $p < .01$, and $M_\Delta = -.50$, $p < .001$, respectively).

### Classification Analysis

A final analysis evaluated which combination of items accurately classified children with and without LI. Item difficulty, which is the percentage of children scoring correctly on a given item, for children with TD and LI was calculated

**Figure 2.** Interaction plots for the significant interactions observed between bilingual experience group (HSE, BESE, HEE) and ability status for copula and passive forms. HSE = high Spanish experience; BESE = balanced English–Spanish experience; HEE = high English experience; LI = language impairment; TD = typically developing.
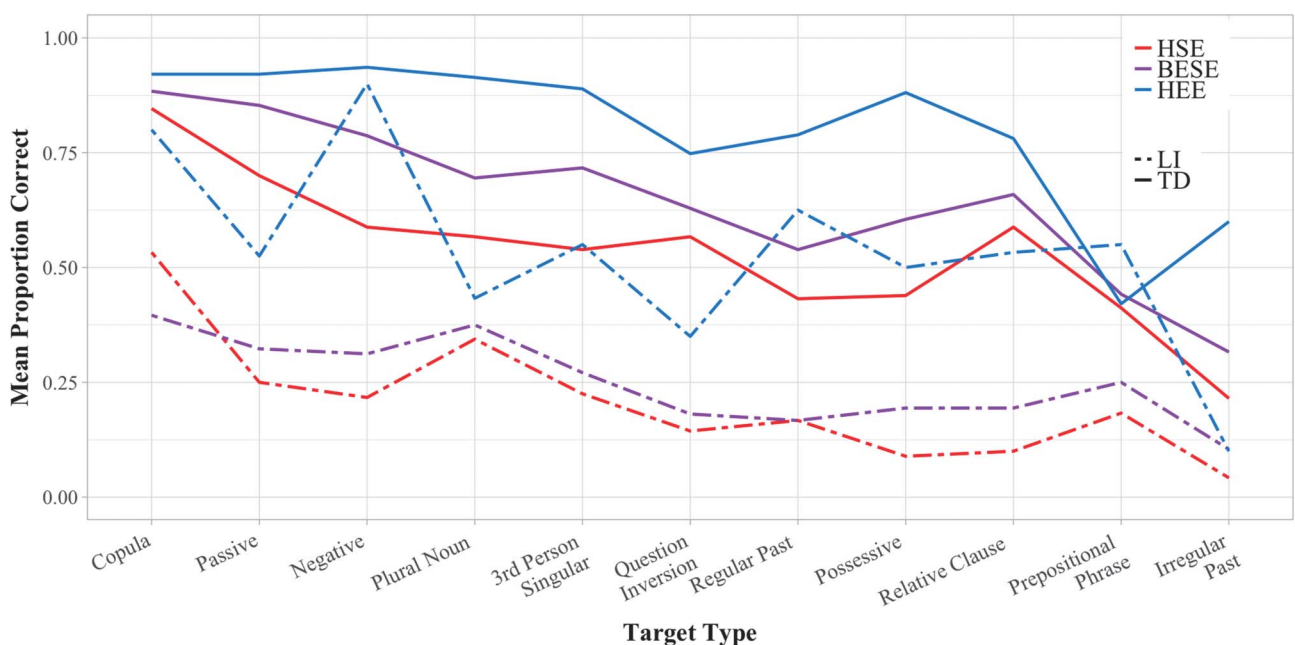
for each of the three bilingual experience groups. Next, differences in item difficulty between children with TD and LI were calculated. We focused on items with difficulty differences of .3 (30%) or more for each of the three bilingual experience groups (Allen & Yen, 2002; Friedenberg, 1995). Table 4 highlights these item types in green. Of the 37 items tested, 19 items met the criteria of reliably differentiating the HEE-LI and HEE-TD groups, 30 met criteria for the BESE-LI and BESE-TD children, and 23 met criteria for the HSE-LI and HSE-TD children. For the purpose of the current analysis, we selected items on the basis of the HEE and BESE groups. Specifically, nine items were selected that met the item discrimination criteria of .3 or greater and where the difficulty levels between the HEE-TD and BESE-TD children were similar (below .2 item difficulty difference). This procedure is intended to minimize differences by these two bilingual experience groups and to maximize differences by ability. We were particularly interested in whether items selected for the HEE and BESE groups would accurately classify the HSE group that was in the process of learning English. A raw score on the basis of the sum items correct of the nine items was calculated for every participant, and these scores were subjected to discriminant analysis.

First, we tested the classification accuracy of the raw score with the BESE and HEE groups pooled. The assumption of equality of the group covariance matrices as tested by the Box's M statistic was met ($p = .080$), and the log determinants were similar (LI = 1.846, TD = 1.401). The chi-square test was significant (Wilks $\lambda = .661$, $\chi^2 = 95.358$, $df = 1$, canonical correlation = .582, $p < .001$). Children with LI had an average raw score of 2.82, and those with TD had an average raw score of 7.05. The total raw score classified 87.1% of the cases accurately with 79.4% sensitivity and 88.4% specificity. We used the same function to classify the HSE children. Classification of the HSE children was 75.2% accurate with 90.3% sensitivity and 71.1% specificity.

Because the cut-points for the HSE group were set to those empirically derived for that of the BESE and HEE groups, we wondered if setting optimum cuts derived from the HSE scores would improve classification. This procedure illustrates the process of local norming (Junker & Stockman, 2002). We reran the discriminant analysis, including only the HSE group. The assumption of equality of the group covariance matrices as tested by the Box's M statistic was not met ($p = .043$). The log determinants were dissimilar (LI = 1.395, TD = 2.027). Thus, we reran the analysis using separate covariance matrices. The results did not change, so we report the original results. The chi-square test was significant (Wilks $\lambda = .696$, $\chi^2 = 51.636$, $df = 1$, canonical correlation = .551, $p < .001$). Children with LI had an average raw score of 1.65, and those with TD had an average raw score of 5.83. The total raw score classified 80.7% of the cases accurately with 83.9% sensitivity and 79.8% specificity.

## Discussion

SLPs need to determine whether children's English errors, such as low accuracy on past tense production or omission of third-person present tense marking, are due to LI or to acquiring English as a second language. As such, we need to systematically profile and identify language disorder within the expected differences in ELLs' performance. This study investigated the role of 11 item types from the experimental version BESA-ME (Peña et al., 2008). We also explored whether differences in exposure, as indexed by bilingual experience group status, led to different test outcomes.

### Grammatical Forms by Bilingual Experience Group and Impairment Status

Overall, the percentage of accuracy of the items tested reflect the attested variability in the language of children who are acquiring English as a second language as illustrated in Table 4. To date, we do not have a convergent picture of expected levels of accuracy for the production of English grammatical accuracy. Some studies report accuracy of specific forms (Blom & Paradis, 2013), whereas other studies report percentages of children who have mastered particular forms (e.g., Dulay & Burt, 1974; Davison & Hammer, 2012). What is important to the approach taken here is that we focused on differences in scores that lead to correct classification rather than absolute accuracy. The finding that many of the forms tested differentiated children with and without LI who use English regularly was not unexpected as grammar is a robust clinical marker of LI in both monolingual and bilingual children (Bedore & Leonard, 1998; Gutierrez-Clellen & Simon-Cereijido, 2007; Rice & Wexler, 1996). When examining the children's total scores for grammatical production, performance did not significantly differ across bilingual experience groups (i.e., HSE, BESE, and HEE), and differences for ability level were observed for the majority of the forms tested for BESE and HEE groups but not for HSE children. Thus, it appears that many of the types of forms tested would contribute to the correct classification of LI for ELL children across all levels of English use. But, at the item level, only a small number of the set tested demonstrated similar item difficulty and were informative for both BESE and HEE. These same items were informative for the HSE group, but their accuracy was lower overall. For children who have less than 40% of English use or for children with less than 6 years of cumulative exposure, it is important to include L1 testing in the assessment plan.

Several forms stood out due to differences in performance patterns. As indicated above, the observed accuracy here differs somewhat from past reports regarding grammatical development in ELLs. Some forms were quite accurate. Copula production showed a group by ability interaction. As illustrated in Figure 2, copula was quite accurate for the TD children. This boosted level of performance is consistent with Paradis' observation of acceleration of acquisition of the copula *be* in bilingual learners (Paradis & Blom, 2016). The copula in Spanish may be more salient because it is used in very similar ways to the English copula, and in Spanish, it is produced as a full word form, whereas in

English, it is often reduced. From a usage-based learning perspective, this may facilitate ELLs treating such an element as a chunk (Bybee, 2008; Tomasello, 2003). The children with HSE-LI produced copula *be* more accurately than the children with BESE-LI. In spite of this appearing to reflect greater knowledge, it may, in fact, reflect an earlier phase of learning where children produce learned chunks (Plunkett & Marchman, 1991). The copula was not an informative predictor for the HEE. The children with HEE-LI did not perform significantly below their typical peers. This pattern likely reflects longer term experience with English, and by 7 to 10 years of age, these children have almost reached mastery of the form.

The negative construction also demonstrated a group by ability interaction. Forms, such as the negative, bear special mention because it a form that is constructed differently in Spanish and English. In Spanish, one only needs to include the negative element (Ella no corre "She no run"), whereas in English, a negative element and the dummy auxiliary, which has no parallel in Spanish ("She does not run"), are required for a negative construction. The differences across groups highlight the difficulty of these non-parallel forms as a function of experience. When children have more extensive experience, they are more accurate in their production of these forms. Here, TD HSE children were less accurate than their BESE peers, who were less accurate than their HEE peers. In contrast, children with HSE-LI and BESE-LI scored below their TD peers and below HEE-LI peers. The negative construction was not informative for HEE children as those with LI and TD were both highly accurate on this construction. This is a good example of the ways that experience will make a difference in the kinds of items that differentiate children with and without LI.

Prepositions also followed a slightly different pattern across groups. Prepositions showed differences by exposure groups and ability level. Taliancich-Klinger et al. (2017) observed English prepositions to be quite difficult for bilingual school-age children, and English production was not predicted by knowledge of Spanish prepositions. For children in the LI HSE and BESE groups, performance is very low. Children with TD are more accurate, but they are still near the 50% accuracy level. Only the HEE-LI and HEE-TD did not show differences. These findings highlight that, when knowledge in the L1 does not map on the L2 system persistence, low performance will result in minimal differences in TD children and in children with LI.

The children's performance on regular and irregular past tense forms also bears mention. These forms contribute to classification accuracy for English-speaking children (e.g., Bedore & Leonard, 1998; Rice & Wexler, 1996). But past tense is challenging for Spanish speakers acquiring English as a second language (e.g., Jacobson & Schwartz, 2005). The TD children in this study produced regular past with 59% accuracy overall and irregular past with 38% accuracy. This level of accuracy for regular past is in line with what has been reported by other works focusing on ELLs (e.g., Blom & Paradis, 2015; Gutierrez-Clellen & Simon-Cerejeido,

2007; Jacobson & Schwartz, 2005). Only Jacobson and Schwartz (2005) have reported irregular past separately, and their participants were more accurate at 47% accuracy. TD and LI differences for HEE groups were not sufficiently robust on regular past to contribute to correct classification, as the differences between the two groups were small. TD and LI differences for BESE and HSE groups may be sufficiently robust to contribute to correct classification, but the means for the TD groups were relatively low, indicating that this form is too difficult. Irregular past was very difficult for all LI groups, with an overall accuracy of 8%. The HEE-TD group was 60% accurate on this form, but the BESE-TD and HSE-TD groups were less than 50% accurate, again indicating that this form is too difficult to be informative.

Overall, the performance patterns on the forms included in the final composite highlight that, even when there are differences in production patterns as a function of ability group, not all forms can be expected to function well as clinical markers. If there is overlap between the lowest performing TD group or variability associated with ELLs performance, then low performance by the children with LI may not help classify children accurately. It is important then to focus on forms that are stable for TD ELLs to most effectively identify ELLs with LI.

### Classification Accuracy for Three Levels of Bilingual Experience: Clinical Implications

To evaluate classification accuracy, we generated a raw score composite on the basis of the nine items that best discriminated impairment for both the HEE and BESE that simultaneously showed similar difficulty levels for the typical children. The raw score composite was subjected to discriminant analysis for the HEE, BESE, and HSE groups. Using the same cut score across the three groups, we found that classification accuracy was acceptable (> 80%) for the HEE and BESE groups but not for the HSE. When we reran the analysis allowing the program to set the cut score on the basis of LI and TD group means within the HSE group, the classification improved. Clinical implications are that, for BESE and HEE children, it is possible that grammatical markers of LI in English are robust enough to differentiate between children with and without LI. Though TD children did make errors on these forms, consistent with Paradis (2016), the children with LI nonetheless scored significantly lower than those with typical development. It is important to note that only a small set of item types (passives and question inversion) showed similar difficulty and acceptable discrimination levels for the TD HEE and BESE groups to retain. Thus, grammatical measures normed for English-only populations may not be sufficiently informative tools for clinical decision making for bilingual children, even those who use English at least 40% of the time and who have at least 6 years of experience with English. This is the case for two related reasons. Tests for monolingual children are more likely to tap tense marking forms that are not informative for children at early stages of L2

acquisition. Further, norms for English language tests do not provide normative data for this population. This finding aligns with other comparisons of grammatical acquisition (e.g., Paradis, 2016) in that morphosyntax has the most prolonged period of acquisition. For the HSE group, these same item types were informative but required a different cut-point for acceptable classification accuracy. We focused on items in common that worked well across the groups. But, for each of the three groups, there were different sets of items related to level of exposure that distinguished LI and TD. Here, we can see that merely renorming tests on another population and resetting a cut score is not sufficient if the items do not show enough of a differentiation for children with and without impairment within the target group. For these children, it is essential that tests on the basis of their level of acquisition and learning-based measures, such as dynamic assessment, inform decisions.

For HSE children, who used Spanish more than 60% of the time, the items that were robust for BESE and HEE did not accurately classify the two ability groups when the same cut-point was used. High sensitivity came at the cost of overidentification. That is, nearly 30% of the children who had typical development were identified with impairment along with 90% of those with LI. At a practical level, this means that HSE children who score in the typical range in English are likely to have typical language development (Kohnert, Windsor, & Yim, 2006). Yet, when tested in English, a score in the impaired range would be uninterpretable. Resetting cut scores does improve the classification, but note that only a small subset of the items tested had sufficient discrimination levels across all three groups.

Examination of the nine items that were selected for the composite score reveals that the average difference score, or D-values, was similar across the three groups (HEE, D-value = .39; BESE, D-value = .49; HSE, D-value = .46). But the difficulty levels (*p*-value) for the TD groups were dissimilar (HEE, *p*-value = .83; BESE, *p*-value = .75; HSE, *p*-value = .65). The greater difficulty level for the HSE group likely contributed to the lower classification rate for this group. For HSE children, it is important to be highly cautious when interpreting results of English language testing that focus on clinical markers on the basis of English. Not all items that work well for differentiating one group will work as well for differentiating TD and LI in other groups.

In some ways, it is not surprising that a measure of English would not work as well for HSE speakers. In fact, it was not anticipated that the measure would work as well as it did. There were grammatical constructions that were indeed robust for this group of children. It may be that a composite of items that do differentiate among HSE children with and without impairment would be a more focused strategy for identification of LI in this population. It may also be useful to consider that items could function to differentiate impairment for different reasons at different levels of experience. For children at higher levels of exposure (BESE, HEE), items may work because children are responsive to the grammatical constructions. For the HSE

children, it is more likely that it is the memory demands of recalling the sequence contributes to the differences between children with and without LI. It would be also important to focus carefully on the common characteristics at the item level.

### Conclusion

For clinicians, grammatical errors in spontaneous conversation and narratives often are "red flags" for concern regarding possible LI. When grammatical errors are considered to be red flags for impairment (Dulay & Burt, 1974), one may be left with the impression that all children who are acquiring English will perform with very low levels of accuracy. Yet, our findings highlight that children who use English more than about 40% of the time can achieve relatively high levels of accuracy in the production of English grammar. However, the variability observed here highlights the need for reconsideration of the cut-points and error patterns for ELLs. This is especially true for the HSE group versus the BESE and HEE groups, which may lead to a more nuanced assessment of risk. In particular, concern regarding clinical markers for English LI, such as irregular past tense, should be considered carefully relative to a child's level of bilingualism.

This study has documented the relationship between bilingual experience and language learning ability in the use of grammatical forms in children acquiring English as a second language. A novel aspect of this study is that discriminant analyses were included as part of the focus and that we considered the effects of three separate bilingual experience groups. It also provides a framework that could be expanded to investigate different languages (e.g., Vietnamese–English bilinguals), different language tasks (e.g., sentences recall), and different language domains (e.g., semantics). A limitation is that a small number of items were tested for each grammatical item type, so more work is needed to broadly replicate the findings. Another area for further consideration is that the participants in this study were young school-age children, but we are often interested in the performance of younger children. Thus, it is important to consider how children with less experience with English will perform on these items and whether the differences in performance of children with TD and LI can be reliably differentiated. Additional work may yield information about items that may help us better understand how LI manifests in the face of the expected variability associated with ELLs and, thus, be able to identify LI in children who are still learning English, such as the HSE bilingual children.

### Acknowledgments

# References

Allen, M. J., & Yen, W. (2002). *Introduction to measurement theory.* Long Grove, IL: Waveland.

Bedore, L. M., & Leonard, L. B. (1998). Specific language impairment and grammatical morphology: A discriminant function analysis. *Journal of Speech, Language, and Hearing Research, 41*(5), 1185–1192.

Bedore, L. M., & Peña, E. D. (2008). Assessment of bilingual children for identification of language impairment: Current findings and implications for practice. *International Journal of Bilingual Education and Bilingualism, 11,* 1–29.

Bedore, L. M., Peña, E. D., Summers, C., Boerger, K., Greene, K., Resendiz, M., & Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual prekindergarten students. *Bilingualism: Language and Cognition, 15*(3), 616–629.

Blom, E., & Paradis, J. (2013). Past tense production by English second language learners with and without language impairment. *Journal of Speech, Language, and Hearing Research, 56*(1), 281–294. https://doi.org/10.1044/1092-4388(2012/11-0112)

Blom, E., & Paradis, J. (2015). Sources of individual differences in the acquisition of tense inflection by English second language learners with and without specific language impairment. *Applied Psycholinguistics, 36*(4), 953–976. https://doi.org/10.1017/S014271641300057X

Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson, N. C. Ellis, P. Robinson, & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216–236). New York, NY: Routledge/Taylor & Francis Group.

Chondrogianni, V., & Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, tense morphology and morpho-syntax in successive bilingual children. *Linguistic Approaches to Bilingualism, 1*(3), 318–345.

Davison, M. D., & Hammer, C. S. (2012). Development of 14 English grammatical morphemes in Spanish–English preschoolers. *Clinical Linguistics & Phonetics, 26*(8), 728–742. https://doi.org/10.3109/02699206.2012.700679

Dulay, H. C., & Burt, M. K. (1974). Natural sequences in child second language acquisition. *Language Learning, 24*(1), 37–53.

Friedenberg, L. (1995). *Psychological testing: Design, analysis, and use.* Needham Heights, MA: Allyn & Bacon.

Gillam, R. B., & Pearson, N. (2004). *Test of Narrative Language.* Austin, TX: Pro-Ed.

Gillam, R. B., Peña, E. D., Bedore, L. M., & Pearson, N. (in development). *Test of Narrative Language–Spanish Adaptation.*

Grela, B. G., & Leonard, L. B. (2000). The influence of argument-structure complexity on the use of auxiliary verbs by children with SLI. *Journal of Speech, Language, and Hearing Research, 43*(5), 1115–1125.

Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics, 24*(2), 267–288.

Gutierrez-Clellen, V. F., & Simon-Cereijido, G. (2007). The discriminant accuracy of a grammatical measure with Latino English-speaking children. *Journal of Speech, Language, and Hearing Research, 50*(4), 968–981. https://doi.org/10.1044/1092-4388(2007/068)

Hammer, C. S., Komaroff, E., Rodriguez, B., Lopez, L., Scarpino, S., & Goldstein, B. G. (2012). Predicting Spanish–English bilingual children's language abilities. *Journal of Speech, Language, and Hearing Research, 55,* 1251–1264.

Hewitt, L. E., Hammer, C. S., Yont, K. M., & Tomblin, J. B. (2005). Language sampling for kindergarten children with and without SLI: Mean length of utterance, IPSYN, and NDW. *Journal of Communication Disorders, 38*(3), 197–213.

Hollingshead, A. A. (1975). *Four-Factor Index of Social Status.* New Haven, CT: Yale University.

Jacobson, P. F., & Schwartz, R. G. (2005). English past tense use in bilingual children with language impairment. *American Journal of Speech-Language Pathology, 14*(4), 313–323.

Johnston, J. R., & Kamhi, A. G. (1984). Syntactic and semantic aspects of the utterances of language-impaired children: The same can be less. *Merrill-Palmer Quarterly, 30*(1), 65–85.

Junker, D. R. A., & Stockman, I. J. (2002). Expressive vocabulary of German–English bilingual toddlers. *American Journal of Speech-Language Pathology, 11*(4), 381–394. https://doi.org/10.1044/1058-0360(2002/042)

Kohnert, K., Windsor, J., & Yim, D. (2006). Do language-based processing tasks separate children with language impairment from typical bilinguals? *Learning Disabilities Research and Practice, 21*(1), 19–29.

Leonard, L. B. (2014). *Children with specific language impairment* (2nd ed.). Boston, MA: MIT Press.

Leonard, L. B., Eyer, J. A., Bedore, L. M., & Grela, B. G. (1997). Three accounts of the grammatical morpheme difficulties of English-speaking children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 40*(4), 741–753.

Lhamon, C., & Gupta, V. (2015, Jan. 7). *Dear colleague letter.* Washington, DC: Department of Justice, Civil Rights Division.

Lugo-Neris, M. J., Peña, E. D., Bedore, L. M., & Gillam, R. B. (2015). Utility of a language screening measure for predicting risk for language impairment in bilinguals. *American Journal of Speech-Language Pathology, 24*(3), 426–437.

Marinis, T., & Chondrogianni, V. (2010). Production of tense marking in successive bilingual children: When do they converge with their monolingual peers? *International Journal of Speech-Language Pathology, 12*(1), 19–28. https://doi.org/10.3109/17549500903434125

Mayer, M. (1967). *A boy, a dog and a frog.* New York, NY: Penguin Books.

Mayer, M. (1969). *Frog, where are you?* New York, NY: Penguin Books.

Mayer, M. (1973). *Frog on his own.* New York, NY: Dial.

Mayer, M. (1974). *Frog goes to dinner.* New York, NY: Penguin books.

Miller, J., & Iglesias, A. (2012). Systematic Analysis of Language Transcripts (Research Version 2012) [Computer software]. Madison, WI: SALT Software.

Newcomer, P., & Hammill, D. (1997). *Test of Language Development–Primary: Third Edition (TOLD-P:3).* Austin, TX: Pro-Ed.

Nicholls, R. J., Eadie, P. A., & Reilly, S. (2011). Monolingual versus multilingual acquisition of English morphology: What can we expect at age 3? *International Journal of Language & Communication Disorders, 46*(4), 449–463. https://doi.org/10.1111/j.1460-6984.2011.00006.x

Paradis, J. (2008). Tense as a clinical marker in English L2 acquisition with language delay/impairment. In E. Gavruseva & B. Haznedar (Eds.), *Current trends in child second language acquisition: A generative perspective* (pp. 337–356). Amsterdam, the Netherlands: John Benjamins.

Paradis, J. (2016). The development of English as a second language with and without specific language impairment: Clinical implications. *Journal of Speech, Language, and Hearing Research, 59*(1), 171–182. https://doi.org/10.1044/2015_JSLHR-L-15-0008

Paradis, J. (2017). Parent report data on input and experience reliably predict bilingual development and this is not trivial. *Bilingualism: Language and Cognition, 20*(1), 27–28. https://doi.org/10.1017/S136672891600033X

Paradis, J., & Blom, E. (2016). Do early successive bilinguals show the English L2 pattern of precocious BE acquisition? *Bilingualism: Language and Cognition, 19*(3), 630–635. https://doi.org/10.1017/S1366728915000267

Paradis, J., Emmerzael, K., & Sorenson Duncan, T. (2010). Assessment of English language learners: Using parent report on first language development. *Journal of Communication Disorders, 43,* 474–497.

Peña, E. D., & Bedore, L. M. (2006). Phenotype assessment tools for bilingual (Spanish–English) children: National Institute of Deafness and Other Communication Disorders.

Peña, E. D., Bedore, L. M., & Gillam, R. B. (2006). Diagnostic markers of language impairment in Spanish–English bilinguals: National Institute on Deafness and Other Communication Disorders.

Peña, E. D., Bedore, L. M., & Griffin, Z. (2010). Cross-language outcomes of typical and atypical development in bilinguals: National Institute of Deafness and Other Communication Disorders.

Peña, E. D., Bedore, L. M., Gutiérrez-Clellen, V. F., Iglesias, A., & Goldstein, B. A. (2006). *Bilingual English–Spanish Assessment—Middle Extension Field Test Version (BESA-ME)*. Unpublished manuscript.

Peña, E. D., Bedore, L. M., Gutiérrez-Clellen, V. F., Iglesias, A., & Goldstein, B. A. (2008). *Bilingual English–Spanish Assessment—Middle Extension Experimental Version (BESA-ME)*. Unpublished manuscript.

Peña, E. D., Bedore, L. M., Gutiérrez-Clellen, V. F., Iglesias, A., & Goldstein, B. A. (2016). *Bilingual English–Spanish Assessment—Middle Extension Field Test Version (BESA-ME)*. Unpublished manuscript.

Peña, E. D., Bedore, L. M., Gutiérrez-Clellen, V. F., Iglesias, A., & Goldstein, B. A. (in development). *Bilingual English–Spanish Oral Screener (BESOS)*.

Peña, E. D., Gutiérrez-Clellen, V. F., Iglesias, A., Goldstein, B. A., & Bedore, L. M. (2018). *Bilingual English–Spanish Assessment (BESA)*. Baltimore, MD: Brookes.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perception: Implications for child language acquisition. *Cognition, 38,* 43–102.

Rice, M. L., Hoffman, L., & Wexler, K. (2009). Judgments of omitted BE and DO in questions as extended finiteness clinical markers of specific language impairment (SLI) to 15 years: A study of growth and asymptote. *Journal of Speech, Language, and Hearing Research, 52*(6), 1417–1433.

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Ryttin, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research, 53,* 333–349.

Rice, M. L., & Wexler, K. (1996). Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech and Hearing Research, 39*(6), 1239–1257.

Rice, M. L., & Wexler, K. (2001). *Rice/Wexler Test of Early Grammatical Impairment*. San Antonio, TX: Harcourt Assessment.

Rice, M. L., Wexler, K., Marquis, J., & Hershberger, S. (2000). Acquisition of irregular past tense by children with specific language impairment. *Journal of Speech, Language, and Hearing Research, 43*(5), 1126–1145.

Seymour, H. N., Roeper, T. W., & de Villiers, J. (2003). *Diagnostic Evaluation of Language Variance*. San Antonio, TX: The Psychological Corporation.

Sullivan, A., & Bal, A. (2013). Disproportionality in special education: Effects of individual and school variables on disability risk. *Exceptional Children, 79*(4), 475–494.

Taliancich-Klinger, C., Bedore, L. M., & Peña, E. D. (2018). Preposition accuracy on a sentence repetition task in school age Spanish–English bilinguals. *Journal of Child Language, 45*(1), 97–119. https://doi.org/10.1017/S0305000917000125

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.

Verhoeven, L., Steenge, J., & van Balkom, H. (2012). Linguistic transfer in bilingual children with specific language impairment. *International Journal of Language & Communication Disorders, 47*(2), 176–183. https://doi.org/10.1111/j.1460-6984.2011.00092.x